Patient Phenotyping Using Interpretable Clustering to Study Clinical Outcomes in TAVR Patients

Roy S. Zawadzki, Terri Johnson, Saman Parvaneh

Edwards Lifesciences, Irvine, USA

Abstract

Discovering and understanding patient profiles, or phenotypes, of the intended population for a treatment is valuable for delivering safe and effective care. This paper explores an interpretable patient phenotyping framework through k-means clustering, statistical tests, visualizations, and decision trees. We use data from a single-center study on transcatheter aortic valve replacement in Germany (n= 581) as an applied example. We clustered the data on continuous demographic, medical history, and pre-operational variables using K-means clustering. We find six distinct clusters and, furthermore, a particular cluster with a statistically significantly higher incidence of myocardial injury during operation compared to the other clusters. Decision tree rules to predict this cluster based on the clustering variable are extracted and compared to the literature for clinical plausibility. Our proof-of-concept analysis highlights the potential of interpretable clustering to understand patient phenotypes. This methodology can be used to find clinically meaningful associations between phenotypes and adverse events.

1. Introduction

Heterogeneity in cardiovascular patient populations may lead to differential outcomes from medical interventions [1]. Therefore, discovering and understanding subpopulations (phenotypes) pertaining to patient safety profiles is valuable to delivering optimal patient care.

Because patients are complex and multifaceted, analyzing patients univariately (e.g., just using age) may yield oversimplification. In this setting, machine learning, specifically unsupervised clustering, can be used to capture complex relationships across a large set of factors. Following this, the clusters can be explored regarding their clinical meaningfulness and association with adverse events.

To illustrate this analysis framework, we focus on a public dataset surrounding the occurrence of adverse events due to transcatheter aortic valve replacement (TAVR).

While similar cluster analyses exist in the settings of aortic stenosis and atrial fibrillation [2, 3], we extend the methods used to evaluate the meaning of the clusters beyond traditional statistical methods by incorporating decision trees.

2. Methods

2.1. Study Cohort

We utilize data that was originally collected and analyzed by Pollari et al., a German single-center study on TAVR indicated for severe stenosis of the native aortic valve [4]. The dataset contains patient demographics, medical history, baseline echocardiography, and pre-operational variables on 581 patients. For more information on the dataset, please refer to the paper. The data is freely accessible online [4] and may be used under the creative commons 4.0 license.

2.2. Dataset Preparation

Our dataset preparation procedure predominately mimicked Kwak et al. [2]. To reduce the number of continuous clustering variables, we first removed variables with a pairwise Pearson correlation on the non-missing data of 0.6 [2]. Then, the more clinically meaningful variable was selected in each pair. For example, between aortic valve delta max and mean, we chose the mean.

Following this, missing data for all the baseline patient characteristics was imputed using the *missForest* algorithm [5]. The continuous variable with the highest missingness was aortic valve delta mean at 41.8% and we had reasonable predictors for all variables, thus, all variables were imputed [6].

2.3. Unsupervised Clustering

After standardizing the variables, k-means clustering was performed on the imputed 17 continuous variables that remained after removing the correlated variables (Table 1).

The optimal number of clusters was selected via three methods: locating the elbow in the sum of squares error, maximizing the Bayesian information criterion (BIC), and

 ISSN: 2325-887X DOI: 10.22489/CinC.2022.345

maximizing the silhouette score. We examined one to 20 clusters.

Table 1. Clustering columns

Column Name	Units
Age	years
Body Mass Index (BMI)	kg/m^2
Creatinine Levels	mg/dl
Creatinine Clearance	mL/min
Euroscore II	n/a
Ejection Fraction	%
Pulmonary Hypertension (PHT)	mmHg
Aortic valve (AV) Delta Mean	mmHg
AV Effective Orifice Area	cm^2
Annulus Area	cm^2
Distance from Annulus to Right Coronary Artery (RCA)	mm
Distance from Annulus to Left CA (LCA)	mm
Valve Oversizing	%
Eccentricity Index	n/a
Total Calcium in AV	mm^3
Total Calcium in Left Ventricular Outflow Tract (LVOT)	mm^3
RCC Calcium in LVOT	mm^3

2.4. Interpreting the Clusters

The clusters were interpreted using both statistical methods and decision trees. First, the differences in the averages of the clustering variables across the clusters were compared using a one-way analysis of variance (ANOVA) and visualized using boxplots. Family-wise error rate was controlled via a Bonferroni adjustment.

A decision tree was fit to predict cluster membership using the clustering variables. To mitigate overfitting and minimize prediction error, we conducted five-fold cross-validated hyperparameter tuning on the tree max depth, minimum samples per leaf, and minimum samples per internal node. Gini feature importance and the decision rules were extracted from this decision tree.

The adverse events of death, major bleeding, and myocardial injury (MI) were combined with the clustering results to view the association with the assigned cluster. There were other adverse events available in the dataset; however, these variables were either not of direct interest or occurred at too small of a rate.

To test differences in event rate per group across clusters, a χ^2 test was conducted. If any results were statistically significant (p-value < 0.05) then follow-up pairwise two-sample proportion tests were run and Bonferroni-adjusted for multiple testing.

For clusters with higher or lower incidence rates of an adverse event compared to the other clusters, we extracted the decision tree rules to arrive at this cluster and search the literature to see if they were clinically plausible. All analyses were performed using Python 3.6 besides the post-hoc two-proportion tests for adverse event rates, which used R version 4.1.1. Decision tree visualization used the package "dtreeviz" [7].

3. Results

Six clusters were found via the elbow method and maximized both BIC and silhouette score. The number of patients in each cluster varied from 13 to 193, and each clustering variable was statistically significant in the one-way ANOVA test (Table 2).

We can visualize the results in Table 3, such as in Figure 1. For example, we observe that cluster 0 has typically a lower ejection fraction than the rest.

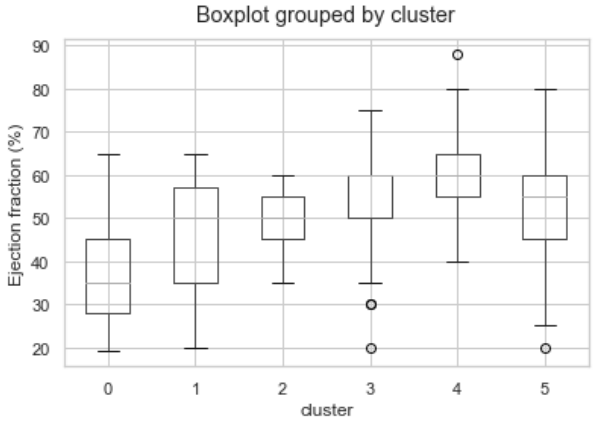

Figure 1. Difference of ejection fraction across clusters.

The optimal hyperparameters for the decision tree were a depth of eight, five minimum samples per leaf, and eight samples for the split. The top ten feature by Gini importance is presented in Table 3. For visualization purposes, a decision tree was fitted with the top four features (Figure 2).

Table 3. Top 10 feature by Gini importance

Variable	Gini Importance
Annulus Area	0.170
Creatinine Clearance	0.169
Aortic Valve Delta Mean	0.111
Ejection Fraction	0.106
Total Calcium in Aortic Valve	0.099
BMI	0.094
Creatinine Levels	0.081
Euroscore II	0.031
Age	0.028
PHT	0.026

Only MI was significantly different across groups in the χ^2 tests. Particularly, cluster 2 had a higher incidence than

the others (Figure 3). Except for cluster 1, likely due to the low sample size in this cluster, there was a statistically significant between cluster 2 and the other clusters ($p < 0.05$).

4. Discussion

In this paper, we have demonstrated a proof-of-concept analysis framework to understand patient phenotypes within populations and their association with clinical outcomes of interest. Communicating findings from unsupervised learning with non-technical audiences and, moreover, generating actionable insights is often challenging. In this vein, our analysis aims to close this gap and, thus, demonstrate the use of clustering techniques for patient phenotyping.

A more traditional approach could be to use a much smaller subset of variables to find vulnerable subpopulations based on statistically significant associations with adverse events; however, there are two main limitations to this process. Firstly, we may inflate multiplicity bias and risk finding spurious relationships by testing against the outcome of interest when choosing which variables to stratify patients upon. This is especially an issue in situations where clinical knowledge is lacking. Secondly, we may trade specificity in identifying patient phenotypes for more interpretability in clinically evaluating these groups.

If clustering is instead used, for the first limitation, the creation of clusters is agnostic to information about adverse events and serves as a data-driven way of differentiating patients. Furthermore, the clusters can be compared across a variety of contexts, not just a singular adverse event. For the second limitation, clustering can handle a greater number of variables, increasing specificity, and, following this, we can integrate our methods for interpretability.

The use of decision trees is advantageous in order to both elucidate the distinction of clusters (e.g., Figure 2) and create clearer rules for determining whether an individual is part of a subpopulation. For example, the 4.5 cm^2 annulus area cut-off, we see this is close to well known the standard definition of small aortic annulus area [8]. We find that within this population, to arrive to the highest incidence of MI cluster (cluster 2), we must also have creatinine clearance of less than 58.47 and a mean gradient of less than 37.78. These cut-offs can potentially serve as hypothesis generating evidence for further study of subpopulations and adverse events.

Table 2: Descriptive statistics of clusters where all numbers are reported as mean (SD). All variables are significant at the 0.003 level (Bonferroni adjustment).

Variable	Overall	Cluster 0 (n = 104)	Cluster 1 (n = 13)	Cluster 2 (n = 25)	Cluster 3 (n = 123)	Cluster 4 (n = 193)	Cluster 5 (n = 123)
Age	81.7 (6.1)	80.6 (5.5)	81.2 (5.5)	74.8 (7.0)	77.5 (6.3)	84.3 (4.6)	84.3 (4.6)
BMI	27.1 (4.8)	25.7 (3.3)	24.7 (3.6)	27.6 (5.1)	31.6 (5.3)	26.2 (4.1)	25.3 (3.4)
Creatinine Levels	1.5 (1.0)	1.5 (0.5)	1.3 (0.4)	5.5 (2.0)	1.2 (0.3)	1.3 (0.4)	1.3 (0.4)
Creatinine Clearance	45.0 (19.6)	41.2 (14.3)	41.7 (14.9)	14.3 (6.3)	67.4 (20.2)	37.6 (12.4)	44.1 (13.0)
Euroscore II	0.1 (0.1)	0.2 (0.1)	0.1 (0.1)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)	0.1 (0.0)
Ejection Fraction	52.6 (12.9)	37.1 (11.5)	46.1 (15.4)	49.3 (8.2)	55.0 (10.6)	59.9 (8.4)	53.2 (10.8)
PHT Measurement	50.6 (13.8)	54.1 (14.9)	49.1 (14.7)	58.3 (15.9)	47.6 (12.8)	50.9 (13.6)	48.6 (12.4)
AV Delta Mean	44.6 (14.7)	30.0 (10.5)	48.1 (14.0)	45.1 (14.0)	42.2 (12.4)	47.6 (12.5)	53.2 (10.8)
AV Effective Orifice Area	0.7 (0.2)	0.8 (0.2)	0.6 (0.2)	0.7 (0.1)	0.8 (0.1)	0.6 (0.1)	0.7 (0.1)
Annulus Area	4.7 (0.9)	4.9 (0.9)	4.6 (1.1)	5.1 (0.8)	4.9 (0.8)	3.9 (0.6)	5.3 (0.8)
Distance from Annulus to RCA	15.6 (3.8)	15.7 (3.7)	14.5 (3.3)	16.4 (3.3)	16.2 (4.0)	14.1 (3.2)	17.1 (4.0)
Distance from Annulus to LCA	13.6 (3.0)	13.6 (2.8)	13.0 (2.6)	14.0 (2.4)	14.4 (3.1)	12.5 (2.6)	14.2 (3.2)
Valve Oversizing	13.8 (15.7)	13.5 (14.1)	14.9 (21.9)	11.4 (11.5)	14.4 (13.9)	20.0 (17.1)	3.9 (11.0)
Eccentricity Index	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)
Total Calcium in AV	839.2 (572.4)	604.6 (418.9)	1529.1 (1167.8)	796.4(443.8)	757.7 (396.4)	596.7 (313.0)	1435 (607.5)
Total Calcium in LVOT	73.2 (123.3)	39.9 (68.2)	431.5 (292.0)	66.0 (77.3)	52.5 (82.9)	53.6 (91.1)	116.2 (145.8)
RCC Calcium in LVOT	5.9 (21.0)	4.2 (9.0)	128.5 (40.4)	1.8 (5.7)	2.0 (7.0)	2.9 (7.8)	3.8 (8.6)

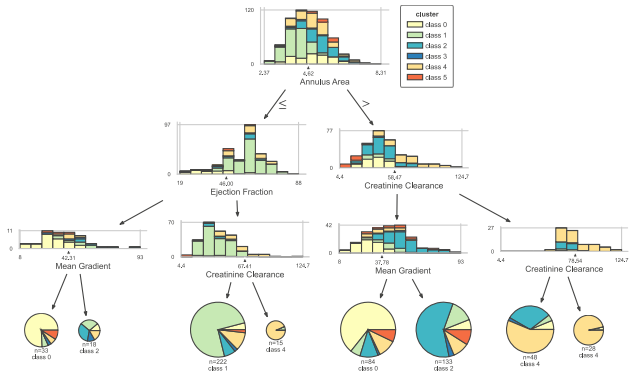

Figure 2. Simplified visualization of the decision tree to predict cluster membership

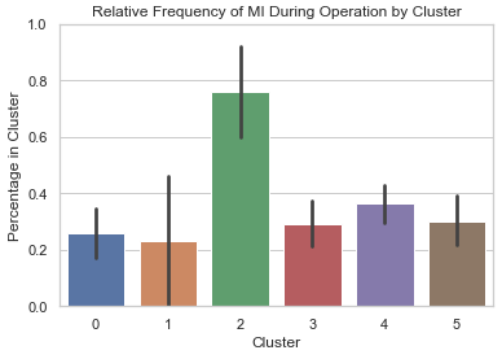

Figure 3: Relative frequency and 95% confidence intervals (black lines) of myocardial injury during operation cluster.

This framework is not without limitations. The number of clusters and patients per cluster may be sensitive to the initial conditions of the clustering algorithm. That is, if one were to take another sample from this same population, the results might change. As the size of such a sample, we would expect this to occur less frequently. Therefore, future studies with a larger population are warranted.

The use of decision trees for interpretability is limited by the fact that lower levels of a decision tree are conditional on rules higher up in the tree. In addition, a decision tree may have multiple leaf nodes for predicting a given cluster. Thus, if the tree complexity is high, visualizing the decision tree will be difficult, and the interpretation of each cluster will be complicated. Future directions could investigate the trade-off between tree fit and interpretability for patient phenotyping. One direction to explain the association between the variables and the cluster could be to extract Shapley additive explanations (SHAP). This approach would be able to accommodate other predictive algorithms and more variables [9].

References

[1] S. M. Manemann, Y. Gerber, S. J. Bielinski, A. M. Chamberlain, K. L. Margolis, S. A. Weston, J. M. Killian, and V. L. Roger, "Recent Trends in Cardiovascular Disease Deaths: A State Specific Perspective," *BMC Public Health*, vol. 21, no. 1, pp. 1-7, 2021.

[2] S. Kwak, Y. Lee, T. Ko, S. Yang, I.-C. Hwang, J.-B. Park, Y. E. Yoon, H.-L. Kim, H.-K. Kim, and Y.-J. Kim, "Unsupervised Cluster Analysis of Patients With Aortic Stenosis Reveals Distinct Population With Different Phenotypes and Outcomes," *Circulation: Cardiovascular Imaging*, vol. 13, no. 5, pp. e009707, 2020.

[3] M. Lachmann, E. Rippen, T. Schuster, E. Xhepa, M. Von Scheidt, G. Harmsen, S. Yuasa, K. Laugwitz, C. Kupatt, and M. Joner, "Sub-phenotyping of Patients With Severe Aortic Stenosis Undergoing Transcatheter Aortic Valve Replacement by Unsupervised Agglomerative Clustering of Echocardiographic and Hemodynamic Data," *European Heart Journal*, vol. 42, no. Supplement_1, pp. ehab724. 1675, 2021.

[4] F. Pollari, W. Hitzl, F. Vogt, M. Cuomo, J. Schwab, C. Söhn, J. M. Kalisnik, C. Langhammer, T. Bertsch, and T. Fischlein, "Aortic Valve Calcification as A Risk Factor for Major Complications and Reduced Survival After Transcatheter Replacement," *Journal of cardiovascular computed tomography*, vol. 14, no. 4, pp. 307-313, 2020.

[5] D. J. Stekhoven, and P. Bühlmann, "MissForest—Non-parametric Missing Value Imputation for Mixed-type Data," *Bioinformatics*, vol. 28, no. 1, pp. 112-118, 2012.

[6] P. Madley-Dowd, R. Hughes, K. Tilling, and J. Heron, "The Proportion of Missing Data Should Not be Used to Guide Decisions on Multiple Imputation," *Journal of clinical epidemiology*, vol. 110, pp. 63-73, 2019.

[7] T. Parr, and P. Grover, "Dtreeviz: Decision tree visualization," GitHub repository, 2020.

[8] A. B. Freitas-Ferraz, G. Tirado-Conte, F. Dagenais, M. Ruel, T. Al-Atassi, E. Dumont, S. Mohammadi, M. Bernier, P. Pibarot, and J. Rodés-Cabau, "Aortic Stenosis and Small Aortic Annulus: Clinical Challenges and Current Therapeutic Alternatives," *Circulation*, vol. 139, no. 23, pp. 2685-2702, 2019.

[9] S. M. Lundberg, and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in neural information processing systems*, vol. 30, 2017.

Address for correspondence:
Saman Parvaneh
1 Edwards Way, Irvine, CA, USA 92614
parvaneh@ieee.org